

сти от начальной подготовки. Кроме того, имеется возможность организации быстрого и эффективного поиска нужных сведений в огромных массивах информации. Традиционно на лекционных занятиях используются такие наглядные средства обучения, как доска и мел, плакаты и схемы, слайды. Эти средства помогали в визуализации информации и в определенной степени повышали качество обучения. Однако подобные формы иллюстративной информации имеют существенные ограничения, так как не способны обеспечить ее восприятие одновременно большим числом студентов. Кроме того, представление такой информации - статично и не воспроизводит динамики процесса, столь свойственных большинству технических дисциплин. Использование цифровых видео- проекторов высокого разрешения, подключенных к компьютеру, позволяет перейти от традиционной технологии, к новой интегрированной образовательной среде, включающей все возможности электронного представления информации. Преподаватель в мультимедиа лекционной аудитории получает вместо доски и мела мощный инструмент для представления информации в разнородной форме (текст, графика, анимация, звук, цифровое видео). В таких системах лектор сам определяет последовательность и формы изложения материала. Мультимедиа лекции в системе высшего профессионального образования можно использовать для преподавания практически всех курсов. Качество и степень освоения учебного материала, как показывает практика, существенно возрастают. Помимо обеспечения богатой образовательной среды, здесь сказывается и то, что преподаватель, сократив время на воспроизведение информации, получает существенно больше времени на объяснение материала.

-
1. "Theory and Practice of Online Learning", Terry Anderson, Fathi Elloumi (eds.), Athabasca University, 2004.
 2. Mark Prensky, Digital Game-Base Learning, McGraw-Hill, 2000.

Хохлов И.А., Емельянова Л.А., Долинская Н.П.,
ПРОГРАММНО-ТЕХНОЛОГИЧЕСКИЙ КОМПЛЕКС ПОДДЕРЖКИ
ПОЛНОТЕКСТОВОЙ ЭЛЕКТРОННОЙ БИБЛИОТЕКИ

hia@uran.ru

ИММ УрО РАН

г. Екатеринбург

Более 10 лет в ИММ УрО РАН выполняется проект по разработке программно – технологического комплекса создания и поддержки электронных библиотек. За это время разрабатываемая система эволюционировала от создания электронных каталогов на магнитных лентах на ЕС ЭВМ до сложной многоуровневой системы создания и сопровождения электронных каталогов в стандарте МАРК доступных для поиска через Интернет. В системе поддерживаются некоторые библиотечные технологии – работа на абонементе, штрихкодирование единиц хранения, большое количество различных справок, необходимых работникам библиотеки. В 2006 г. в систему добавлена возможность работы с полнотекстовыми файлами. Ведется работа по поддержке протокола Z39.50. Поскольку с самого начала проект был нацелен на

использования в научных библиотеках УрО РАН он называется «Система электронных библиотек (СЭБ) УрО РАН»

Одним из основных принципов, которого всегда придерживались разработчики при создании этой системы – система должна распространяться свободно, бесплатно и в исходных кодах, работать в среде бесплатного системного программного обеспечения и разрабатываться с помощью бесплатных инструментальных средств. Такой подход позволит создавать электронные библиотеки в небольших библиотеках, бюджет которых, как известно, существенно ограничен.

Результатом данной работы явилось то, что в нескольких научных библиотеках институтов УрО РАН используется эта система для создания электронных каталогов с доступом через Интернет. Известны случаи использования ее в других организациях, в том числе и в странах СНГ.

В докладе рассматриваются основные возможности и особенности системы электронных библиотек (СЭБ) как свободно распространяемой системы построения электронных научных архивов на основе Web-интерфейса.

Система электронных библиотек разработана для создания тематических баз данных каталожных карточек по различным отраслям знаний в организациях Уральского отделения РАН. Каждая запись БД является библиографическим описанием печатного издания (книги, журнала, статьи и пр.), содержащегося в библиотеке. Запись в базе данных состоит из полей, и может выводиться по запросу пользователя либо в виде аналога каталожной карточки, либо в виде подробного библиографического описания. Список полей является подмножеством формата "РУС-МАРК". Каждая Научная библиотека выбирает поля для своего каталога самостоятельно.

В системе существует возможность поиска из Интернет. В настоящий момент поиск можно проводить в рамках одной библиотеки, но как указывалось выше ведется работа по поддержке протокола Z39.50, с помощью которого можно организовать поиск по одному запросу в разных библиотеках. Все библиотечные сервера научных институтов УрО РАН независимы. На сервере разработчиков системы <http://lib.uran.ru> расположена страница содержащая краткую информацию о системе, ссылки на полную информацию, ссылки на документацию и инструкции, каталог ссылок на библиотеки УрО РАН. На этом же сервере ссылки для скачивания текстов программ системы и инструкций по их установке и эксплуатации. Скачивание системы возможно только после регистрации на сервере.

Каждая библиотека, включенная в систему, расположена на реальном или виртуальном WWW-сервере - специализированном библиотечном сервере (СБС), который может располагаться на сервере Президиума УрО РАН, сервере института УрО РАН или региональных серверах в научных центрах УрО РАН или любом другом сервере Интернет.

Каждый специализированный библиотечный сервер (СБС) может иметь одну или несколько электронных библиотек, которые входят в СЭБ и содержит информационно - поисковые системы (ИПС) и может содержать дополнительные WEB-страницы и подсистемы, определяемые задачами конкретной библиотеки.

Каждый СБС, расположенный на реальном WWW-сервере администрируется работниками института, эксплуатирующими этот сервер и имеющими квалификацию системного администратора ОС FreeBSD, СУБД MY SQL.

СБС, расположенные на сервере Президиума УрО РАН или региональных серверах, поддерживаются системными администраторами этих серверов. В следующих версиях предполагается разработка следующих структур данных и соответствующих ИПС:

- списков печатных трудов научных сотрудников;
- полнотекстовых баз данных;
- баз данных патентной информации;
- каталог изданий на CD ROM или DVD дисках;
- каталог изданий на микро пленках и микрофишах;
- каталог аудио - видео материалов;
- электронных изданий УрО РАН;
- каталог ссылок Интернет.

Более подробно рассматривается функционирование системы на примере ее работы в научной библиотеке института математики и механики УрО РАН. Система обеспечивает систематизацию единиц хранения в тематических базах данных о научных изданиях, собранных в научной библиотеке за 50 лет ее существования.

СЭБ предоставляет удобный механизм для создания архивов научных знаний с большим разнообразием информационных ресурсов: книги, диссертации, препринты, проблемно-ориентированные базы данных по разделам математических наук, мемориальные библиотеки, труды сотрудников Института. В данный момент в электронной библиотеке находятся данные о 75 600 документах, из которых 27 500 составляют книги.

Несмотря на растущую важность Интернета, большая часть знаний в виде культурного и образовательного материала уже существующего сегодня, всё ещё доступна лишь в бумажной форме. Перевод этого информационного богатства в цифровую форму, в виде максимально близком к оригиналу, с достаточно простым доступом и возможностью поиска по текстам, является насущной потребностью настоящего времени. Такую задачу ставят сейчас ведущие мировые библиотеки и данный проект должен участвовать в ее реализации.

В 2006 году изучен мировой опыт и проведен анализ форматов хранения полнотекстовых данных для электронных библиотек и разработаны принципы организации системы хранения полнотекстовых данных связанных с данными структуры электронного каталога системы электронных библиотек УрО РАН. Для разрабатываемой системы выбран формат файлов DjVu. В процессе исследования форматов хранения полнотекстовых документов были исследованы все графические форматы хранения текстовой информации (TIFF, PDF, DjVu, JPEG и др.) и сделан вывод о преимуществах хранения текстов в формате DjVu.

DjVu - это технология сжатия, файловый формат и универсальная платформа специально спроектированная для создания цифровых библиотек на основе печатного, сканированного или цифрового материала. Так, например, для сканированного документа, размер DjVu файла обычно в 3-10 раз меньше чем чёрно-белый TIFF или PDF и в 5-10 чем цветной JPEG. Для реализации хранения полных текстов документов сделано следующее: разработано программное обеспечение модуля административной подсистемы библиотеки, отвечающий за сохранение на WEB-сервере файла

с полным текстом. Структура базы данных электронного каталога модифицирована для размещения ссылки на файл с полный текст документа в формате DjVu. Разработан ряд сервисных программ для обработки отсканированных фалов и сопровождения их в файловой системе WEB- сервера. Разработан WEB- интерфейс просмотра книг в виде файлов “ DjVu”

Разработана и внедрена в библиотеке ИММ УрО РАН технология сканирования книг и преобразования графических файлов в формат DjVu. Создано программное обеспечение для модификации DjVu – файлов с целью просмотра их в виде двухоконного интерфейса (развернутая книга) и внедрения оглавления в DjVu – файл.

К настоящему времени разработаны основные принципы организации подсистемы поиска информации по текстам на основе текстового слоя в файлах DjVu. Графический или, как его называют, передний слой файла DjVu может быть подвергнут распознаванию с помощью системы OCR, а результат внедрён назад в файл DjVu как доступный для поиска уровень "скрытого текста". Существует и соответствующий инструментарий для извлечения такого текста и перевода его в ряд форматов, которые включают каждое слово с координатами его ограничивающего прямоугольника на странице.

В настоящее время получен только первый опыт на пути создания полнотекстовой электронной библиотеки. Отсканировали первые 50 книг, обработали, распознали, поставили ссылку на полный текст книг с библиографических карточек. Наиболее медленный процесс – сканирование книги с помощью планшетных сканеров. Первая сотня ссылок на полный текст книги размещена в электронном каталоге научной библиотеке ИММ УрО РАН.

Сканирование выполнялось на планшетном и потоковом сканерах, документы сохранены в формате tiff , собраны в DjVu-формат сначала без OCR-слоя. Распознавание производилось FineReader 8.0.* С помощью программ болгарского программиста Генчо получены DjVu-файлы с внедренным текстом. По таким DjVu-файлам возможен поиск информации по ключевым словам и перенос выбранного текстового блока в любой текстовый редактор для обработки.

Сканирование на потоковом сканере осуществлялось отдельно для четных и для нечетных страниц в первом варианте сразу из DjVu редактора. Получались два многостраничных файла. В дальнейшем страницы в DjVu редакторе расставлялись в нужном порядке. Во втором случае сканирование велось из Fine Reader и страницы сохранялись с нужной нумерацией, открывались блоком в DjVu редакторе и перекодировались в один DjVu-файл.

Как указывалось выше, наиболее медленный процесс в создании полнотекстовых библиотек – сканирование книги с помощью планшетных сканеров. Поэтому в ИММ УрО РАН разработан проект центра коллективного пользования для институтов УрО РАН по сканированию книг на базе мощного пакетного книжного сканера Эллар ПланСкан с дополнительным программным модулем PDF-BCS2, который будет функционировать с 2007 г. Сканер закуплен на бюджетные средства Региональной целевой программы по развитию вычислительных, телекоммуникационных и информационных ресурсов УрО РАН. Стоимость сканера – 540 тыс. руб.

Другой путь пополнения полнотекстовых библиотек – поиск в Интернет размещенных полнотекстовых изданий и установление ссылки на них или перекачка их на свой сервер, при условии соблюдения авторских прав.

Как ни сложен и долог процесс перевода важнейшего слоя мировой культуры – бумажных носителей информации в цифровой вид, этого не избежать. Авторы проекта надеются, что их скромный вклад в решение этой задачи будет полезен и уже сейчас малобюджетные библиотеки могут наиболее ценные и раритетные экземпляры тестовых документов из своих хранилищ переводить в цифровой вид и сделать их доступными через Интернет всему миру.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Научно-методическая поддержка разработки научных электронных библиотек / С.И. Акимов [и др.] // Электронные библиотеки. 2005. Т. 8, вып. 1.
2. Куйбышев, Л.А Политика ЕС в области оцифровки культурного и научного наследия. Проекты MINERVA и MINERVA PLUS / Л.А.Куйбышев, Н.В.Браккер // Электронные библиотеки. 2005. Т. 8, вып. 3.
3. Лонги, М. «Флорентийская повестка дня»: инициатива по сохранению цифрового наследия / М. Лонги // Электронные библиотеки. 2006. Т. 9, вып. 1.
4. Алгулиев, Р. М. Электронные издательские технологии. Современное состояние и перспективы развития / Р. М. Алгулиев, И. М. Алиев, М. Г. Шарифов // Информатика и образование. 2005. № 6. С.53-59